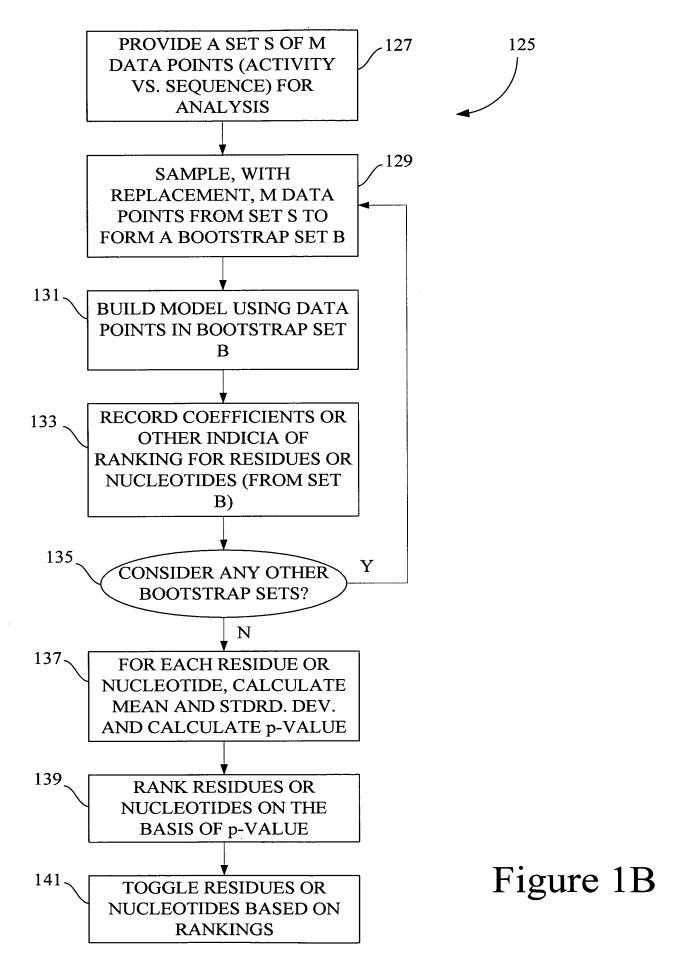
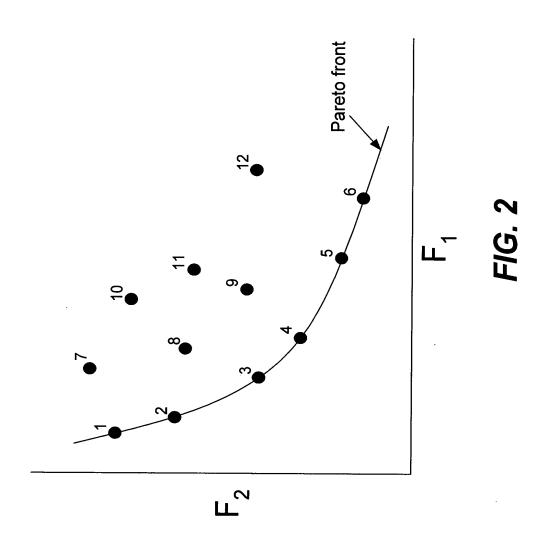
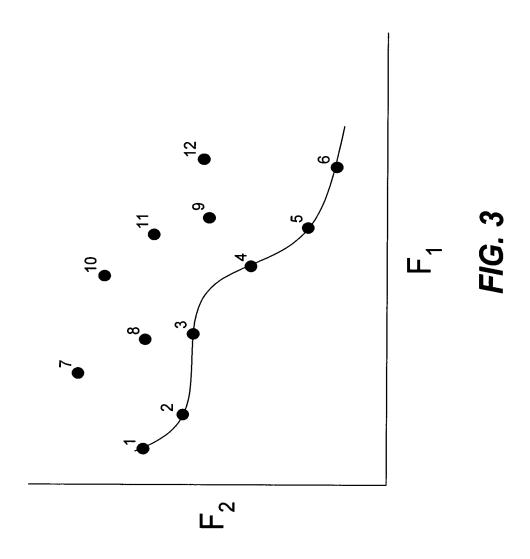


Figure 1A







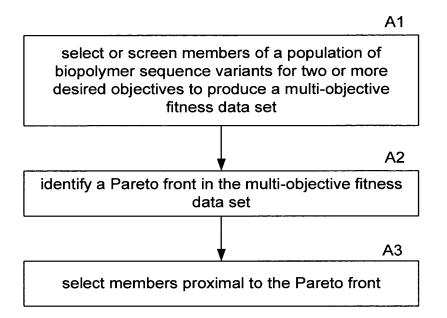


FIG. 4

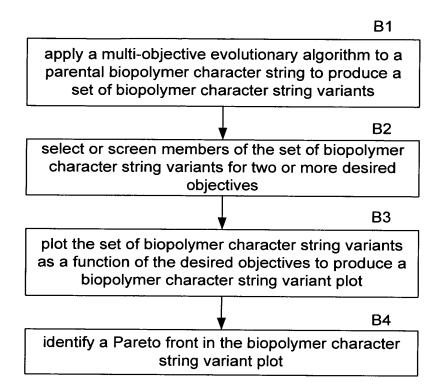


FIG. 5

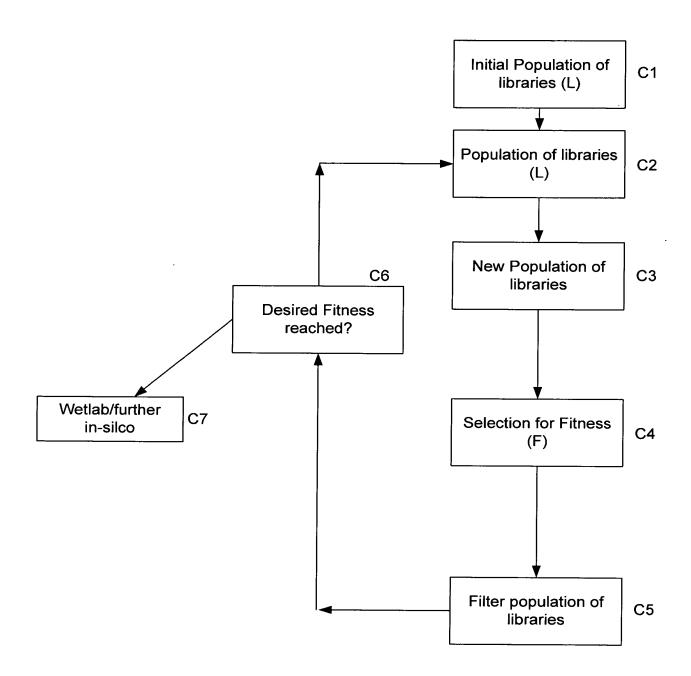


FIG. 6

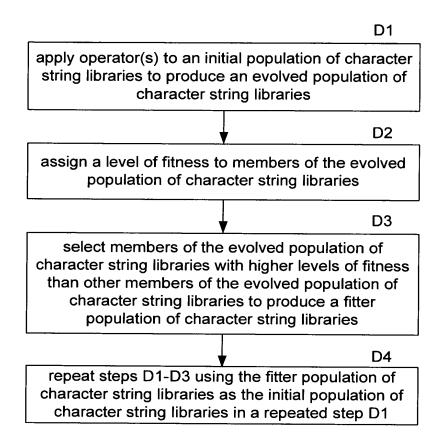


FIG. 7

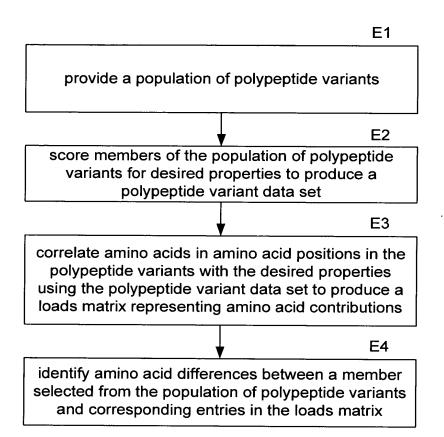


FIG. 8

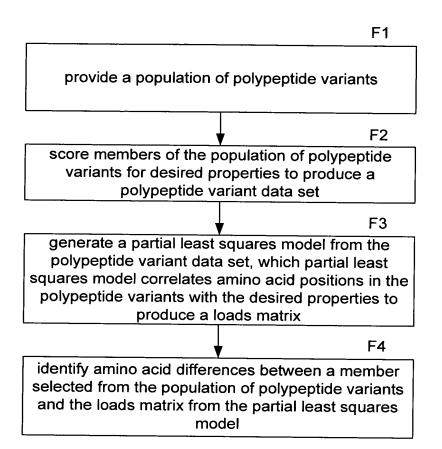


FIG. 9

G1

provide an X predictor matrix that includes a data set corresponding to a set of polypeptide sequence variants in which at least a subset of the set of polypeptide sequence variants include measured activities

G2

calculate cross product terms between or among columns of the X predictor matrix in which each column entry corresponds to an amino acid of a polypeptide sequence variant from the set of polypeptide sequence variants

G3

add a cross product term calculated in G2 to one or more linear terms of the X predictor matrix to produce an expanded X predictor matrix

G4

generate a model with the expanded X predictor matrix to identify important cross product terms and/ or linear terms

FIG. 10

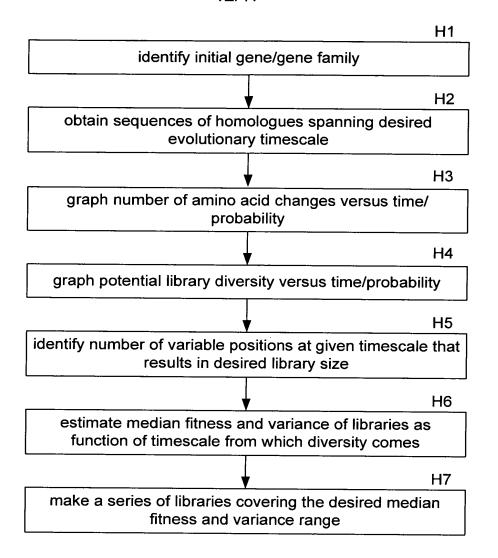


FIG. 11

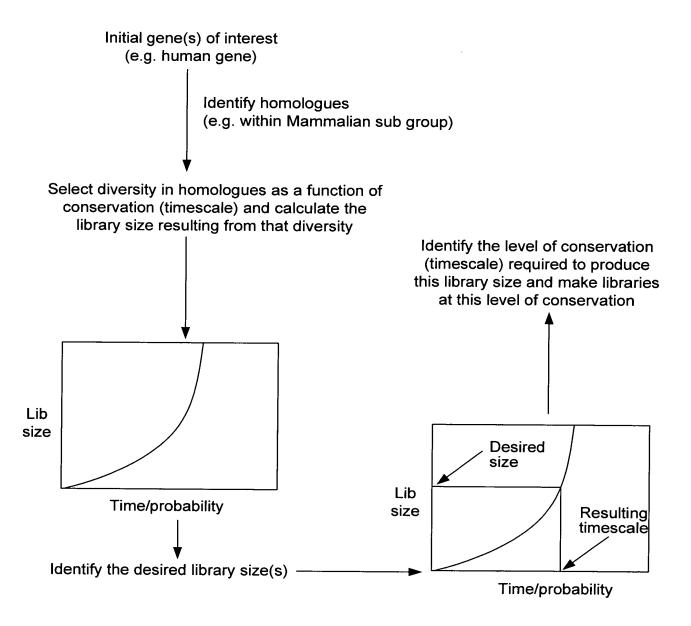
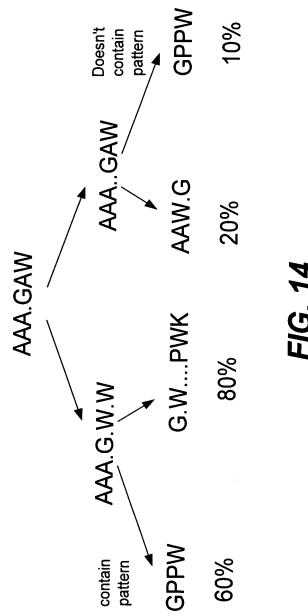


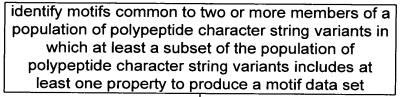
FIG. 12

FIG. 13

predict one or more character strings that include the desired property using the trained neural network



**J1** 



J2

correlate at least one motif from the motif data set with the at least one property to produce a motif scoring function

J3

score at least one target polypeptide character string using the motif scoring function to predict the at least one property of the at least one target polypeptide character string

FIG. 15

